Author(s): Shariatmadari, Hamidreza & Osti, Prajwal & Iraji, Sassan & Jäntti, Riku

Title: Data Aggregation in Capillary Networks for Machine-to-Machine Communications

Year: 2015

Version: Post print

Please cite the original version:

Shariatmadari, Hamidreza & Osti, Prajwal & Iraji, Sassan & Jäntti, Riku. 2015. Data Aggregation in Capillary Networks for Machine-to-Machine Communications. 2015 IEEE 26th International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC): Workshop on M2M Communications: Challenges, Solutions and Applications. 2277-2282.

Data Aggregation in Capillary Networks for Machine-to-Machine Communications

Hamidreza Shariatmadari, Prajwal Osti, Sassan Iraji and Riku Jäntti
Department of Communications and Networking,
Aalto University, P. O. Box 13000, FIN-00076 Aalto, Finland
{firstname.lastname}@aalto.fi

***Abstract*—As machine-to-machine applications using cellular systems become pervasive, it is an important concern that their deployment does not jeopardize the performance of the cellular systems. Support for a massive number of machines brings technical challenges affecting the performance of the random access channel and efficiency of radio resource allocation. Capillary networks are considered as an extensions to the cellular systems for providing large-scale connectivity. This paper proposes an aggregation scheme for capillary networks connected to the LTE network to improve their communication efficiency. A gateway, an intermediate unit between machines and the base station, aggregates packets from the machines during a predefined time, and then delivers them to the LTE network. In addition, this paper analyzes the trade-offs between random access interaction, resource allocation, and communication latency. Results reveals that accepting the extra latency for accumulating packets can significantly reduce the random access requests and the required resources for the data transmissions.**

***Index Terms*—M2M, LTE, Capillary network, Wireless communications**

I. Introduction

Machine-to-machine (M2M) communication or machine-type communication (MTC) is the necessary infrastructure to provide communications between machines without the need for direct human intervention. The interconnection of machines, with the aid of the underlying network infrastructure, allows offering new services and applications in various industries, such as automation, tracking, smart grids, metering, security and public safety [1]. However, these applications have inherently different service requirements compared to the traditional voice and data traffic, which makes the network design more complex.

Cellular systems are a natural choice of technology to provide connectivity for M2M applications due to their wide deployment. Although these cellular systems have mainly been developed and optimized to handle the traffic from human-to-human (H2H) communications, some M2M applications have already been deployed in the existing systems [2]. Long Term Evolution (LTE) system is the new generation of cellular systems, that supersedes the legacy systems and is being deployed widely. Cellular system operators and standardization forums have launched various activities to facilitate M2M communications in this system [3].

M2M applications generally have different characteristics and requirements compared to H2H services. The traffic load in most of the H2H applications is in the downlink, while in M2M applications, the traffic load is usually uplink centric. Moreover, many M2M applications, such as monitoring systems, smart metering, tracking, and vehicular communications, are supposed to support a large number of machines. Introducing a massive number of machines to the LTE system may pose technical challenges concerning the performance of the random access (RA) process, which is used to establish links between the devices and the network, as well as in the efficient utilization of the radio resources.

It has been found that the LTE system is prone to congestion when a large number of devices attempt to connect to the network simultaneously [4]. Various enhancements have been proposed to alleviate this problem. For example, additional resources can be dedicated in the case of congestion [5]. Backoff adjustment schemes and prioritized random access mechanisms are other solutions to keep the congestion to a tolerable level [6], [7]. In order to reduce the transmission delay caused by congestion in link establishment, a packet aggregation method is suggested in which each machine tries to establish a link with the network only when its buffer is filled with a specified number of packets [8]. The number of packets as the threshold can be selected to minimize the overall transmission latency, considering the aggregation period and delivering the data.

This paper proposes a packet aggregation scheme for the capillary networks within the LTE system. Capillary networks are utilized to connect non-LTE devices to the LTE network. They can also be used to group LTE-based devices in order to improve their communications efficiency. In this method, the gateway can aggregate packets from different devices during a predefined period, and then send the aggregated data to the network. We observe that the use of a fixed aggregation period decreases the number of interactions with the network for link establishment, and further reduces radio resources needed for delivering data to the network. Obviously, these improvements are achieved by increasing the overall transmission latency.

The rest of the paper is organized as follows: Section II describes the considered network and the packet aggregation scheme. Section III analyzes the performance of the proposed packet aggregation scheme in terms of the RA process, radio resource allocation, and average communication latency. Finally, the conclusions of the study are provided in Section IV.

II. System model and aggregation scheme

Different network designs and enhancements have been proposed to fulfill the requirements of M2M applications in the LTE system [9]. Gateways, as extensions to the basic LTE architecture, allow statically deployed devices to be connected to the LTE network. As shown in Fig.1, a gateway is at the core of a capillary network within the LTE network and acts as an intermediate entity between the devices and a base station (i.e. eNodeB) by exchanging the data between them. In the downlink, the gateway receives data from the eNodeB, processes it and then delivers it to the appropriate devices. In the uplink, the gateway receives packets from devices and forwards them to the LTE network.

In order to connect a gateway to the LTE network, the link interface between the gateway and the base station should remain in compliance with the LTE radio interface. However, the communication interface between devices and the gateway can be based on the LTE standard, or any other wired or wireless technology. Currently, short-range wireless technologies, such as IEEE 802.15.4 and Bluetooth low energy, are preferred due to the ease of device installation and the low cost of these devices. In addition, the power consumption of these short-range technologies is very low, which enables the devices operate with batteries for a long period.

In this scenario, which is shown in Fig. 1, N devices are connected to the LTE network through a single gateway. Each device attempts to transmit a packet to the gateway whenever it has some data, and the gateway sends an acknowledgment (ACK) message to the device after each successful reception of a packet. However, due to the noise, interference, and collision, the gateway might not receive a packet successfully; hence ACK is not sent. For generality, a special type of communication technology is not considered between the devices and the gateway. Instead, it is presupposed that the communication technology directly affects the link reliability. Therefore, the quality of link is modeled by the probability of packet failure, i.e. the probability that the gateway fails to receive a packet successfully. The packet failure probabilities for all the devices are considered to be the same, do not change over time, and are equal to p_1. In addition, it is assumed that the ACK messages are delivered to devices without any failure. If a device does not receive the ACK message, it considers that the message was not received successfully and tries to retransmit the packet. The maximum number of transmissions attempts for each packet is limited to m times before the packet is dropped.

The gateway can forward each received packet immediately upon reception, or it can aggregate packets from multiple devices and then transmit them. If the number of connected devices to the gateway is high, or the connected devices generate a large number of packets, forwarding packets immediately causes high traffic and signaling load on the LTE network. In order to improve the communication efficiency, we propose a packet aggregation scheme, in which the gateway aggregates packets arriving during a period of time, called the aggregation period, and then forwards all the aggregated packets at once to the base station. The aggregation period, which is D (seconds), begins when there is not any aggregated packet in the gateway and a new packet is successfully received by the gateway. When the gateway intends to transmit data to the LTE network, it should establish a link with the eNodeB in order to send scheduling request (SR) and obtain the required radio resources for data transmissions. The RA process is performed to establish the link. This process might be unsuccessful as multiple devices may try simultaneously to access a shared channel, known as physical random access channel (PRACH). It is assumed that the probability of failure in the RA process is constant over time and is equal to p_2. In the case of a failure, the gateway again starts the RA process after a backoff interval. However, the maximum number of RA attempts for each data transmission is limited to n times in order to avoid network congestion. When the RA is performed successfully and SR is sent, the network provides dedicated resources for the gateway to transmit data. Since the data transmission occurs over a dedicated channel, it can be assumed to be error-free. The process of link establishment and data transmissions are further described in [10].

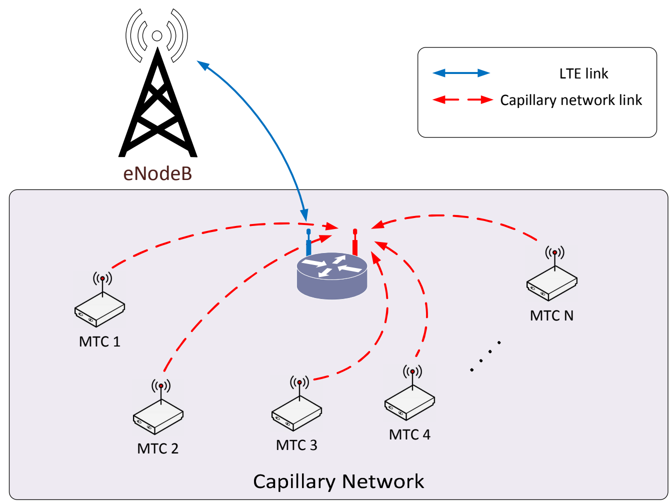

Fig. 1. Considered system model.

Fig. 2 illustrates the packet aggregation scheme, in which the packets from the devices are aggregated, and forwarded to the eNodeB after the end of the aggregation period. Packets may be received by the gateway with different latencies due to the employed medium access control (MAC) protocol and the number of packet retransmissions. Note that Fig. 2 only presents the packets that the gateway could receive successfully while the dropped packets are not shown. At the end of the aggregation period, the gateway starts the RA process to establish a link and forward the data. Transmission latencies may also vary for the aggregated packets originating from the gateway, as the radio resources necessary for the RA process may be available only in certain specific subframes.

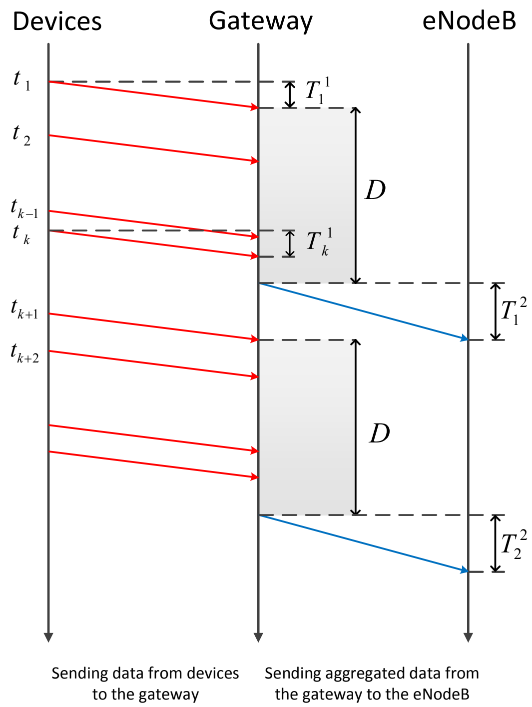

Fig. 2. Packet aggregation scheme.

In addition, if the failure occurs in the RA, the gateway needs to repeat RA again. If the aggregation period (D) shrinks to zero, the gateway immediately tries to forward each received packet. This is similar to the scenario in which the aggregation is not performed.

III. Performance analysis and simulation results

This section provides the performance analysis of the aforementioned packet aggregation scheme in terms of the RA interactions, the radio resource utility, and the average communication latency for delivering packets. To gain insights into the effects of the aggregation scheme in the system, each performance metric is compared when the aggregation method is applied to the non-utilized aggregation case. In addition, closed form formulas for a particular scenario are established and compared with the simulation results.

Many M2M applications, such as sensor and monitoring applications, are formed with a large number of devices, in which devices generate random packets with relatively the same size. Generally, traffic load from each device is low or moderate, however, the overall traffic from all of the devices might become high. In this scenario, it is assumed that N connected devices randomly generate packets, and try to send their data to the LTE network through a single gateway. The size of packets is of a constant value, s bytes, and each device generates packets according to a Poisson process with an average of λ packet arrivals per second. The gateway aggregates packets during the aggregation period D, and then initiates the link establishment through the RA procedure. When the link is established, the gateway transmits the aggregated packets using the radio resources granted during the RA procedure.

A. Performance analysis of RA process

If the aggregation is not applied in the system, the gateway needs to perform the RA process for each successfully received packet from devices separately. The required number of RA processes depends on the connected devices, and the traffic load from them. Packet aggregation decreases the RA interactions. In order to assess the performance of this aggregation method, the normalized required number of RA processes can be defined as the ratio of required RA processes when the packet aggregation method is applied to the required RA processes when the aggregation method is not applied. This can be expressed as:

$$r(D) = \frac{Required\ RA\ processes\ with\ aggregation}{Required\ RA\ processes\ without\ aggregation}.$$

In the aforementioned scenario, on average the gateway receives $N\lambda(1-p_1^m)$ packets/sec successfully. If the aggregation method is not applied, the gateway needs to perform the RA procedure for each received packet. Considering failures in the RA procedure, the gateway performs $N\lambda(1-p_1^m)\sum_{i=1}^{n} ip_2^{i-1}(1-p_2)$ RA processes/sec on the average. When the aggregation method is applied, only one RA procedure is performed after the aggregation period. The average time between two aggregation periods is $D+\frac{1}{N\lambda(1-p_1^m)}$. Hence, on average the gateway needs to perform $\frac{\sum_{i=1}^{n} ip_2^{i-1}(1-p_2)}{D+\frac{1}{N\lambda(1-p_1^m)}}$ random access requests per second. The normalized required RA procedures can be expressed as

$$\begin{aligned} r(D) &= \frac{\frac{\sum_{i=1}^{n} ip_2^{i-1}(1-p_2)}{D+\frac{1}{N\lambda(1-p_1^m)}}}{N\lambda(1-p_1^m)\sum_{i=1}^{n} ip_2^{i-1}(1-p_2)} \\ &= \frac{1}{N\lambda D(1-p_1^m)+1}. \end{aligned}$$

Fig. 3 shows the simulation and analytical results of normalized required RA interactions when 50 devices are attempting to send packets to the network, with two different packet arrival rates. It is assumed that the $p_1 = 0.1$, $p_2 = 0.1$, $n = 5$, $m = 5$. It is evident that the packet aggregation scheme can significantly reduced the RA interactions. The reduction is greater for the higher traffic load; hence, even a small aggregation period greatly reduces the costly RA procedure.

B. Performance analysis of radio resources utility

When the gateway establishes a link with the eNodeB, it performs SR to obtain radio resources for data transmission. The network provides dedicated radio resources and informs the gateway to communicate using those resources. The smallest element of resource allocation in LTE is called a resource block (RB) which corresponds 12 consecutive subcarriers during one slot period. The amount of information

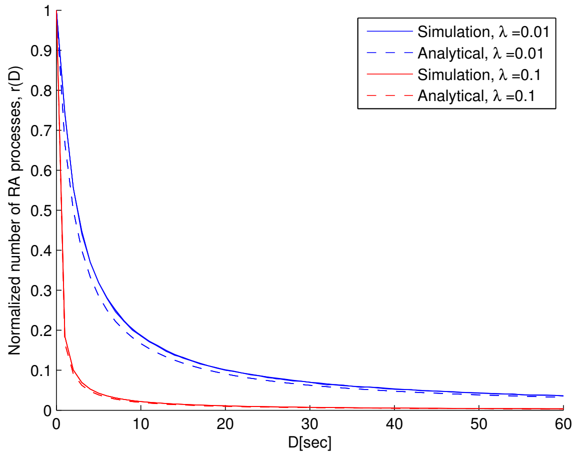

Fig. 3. Simulation and analytical results for normalized required RA procedures.

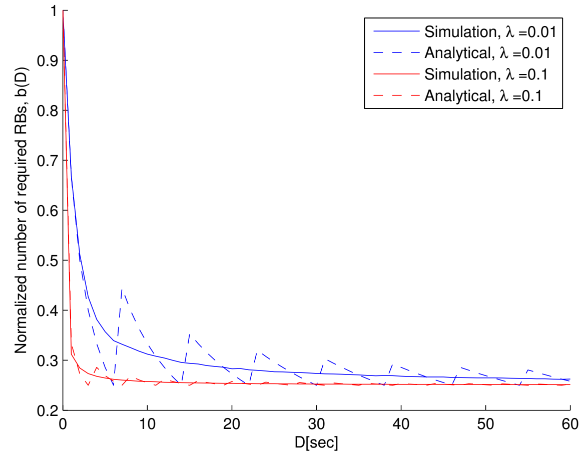

Fig. 4. Simulation and analytical results for normalized required RBs.

can be loaded in each RB depends on the modulation type, coding scheme, signaling overhead, etc. Hence, the number of required RBs for the gateway to transmit data depends on the message size and the amount of information can be fit in one RB. The packet size for many M2M applications is smaller than the capacity of a RB; hence, using a single RB for carrying each packet results in waste of radio resources. The packet aggregation scheme can also reduce the number of required RBs to convey data to the eNodeB by fitting many packets from devices into one RB. To compare the efficiency of data aggregation in terms of RB utility, we define the normalized required RBs as the ratio of required RBs when data aggregation is not applied, to required RBs when the packet aggregation is utilized, i.e.:

$$b(D) = \frac{Required\ RBs\ with\ aggregation}{Required\ RBs\ without\ aggregation}.$$

In the considered traffic scenario, it is assumed that the size of packets generated by devices is s bytes and each RB can carry c bytes of information. If the aggregation method is not applied, $\lceil \frac{s}{c} \rceil$ RBs are utilized for each packet transmission. On the average, the gateway should be granted by $N\lambda(1-p_1^m)(1-p_2^n)\lceil \frac{s}{c} \rceil$ RBs/sec to deliver received packets from devices. When the aggregation is applied, it can be expected that A additional packets arrive during the aggregation period where A is a poisson distributed random variable with rate $(N\lambda D(1-p_1^m))$. If the gateway is able to establish the link with the eNodeB, an average of $\mathbf{E}[\lceil (1+A)\frac{s}{c} \rceil]$ RBs are required for transmitting the aggregated packets at the end of the aggregation period and consequently the gateway requires $\frac{1-p_2^n}{D+\frac{1}{N\lambda(1-p_1^m)}}\mathbf{E}[\lceil (1+A)\frac{s}{c} \rceil]$ RBs/sec in average. The normalized required RBs can then be expressed as:

$$\begin{aligned} b(D) &= \frac{\frac{1-p_2^n}{D+\frac{1}{N\lambda(1-p_1^m)}}\mathbf{E}[\lceil (1+A)\frac{s}{c} \rceil]}{N\lambda(1-p_1^m)(1-p_2^n)\lceil \frac{s}{c} \rceil} \\ &\approx \frac{1}{1+N\lambda D(1-p_1^m)} \frac{\left\lceil \frac{[1+N\lambda D(1-p_1^m)]s}{c} \right\rceil}{\lceil \frac{s}{c} \rceil}. \end{aligned}$$

Fig. 4 shows the simulation and analytical results of normalized required RBs when 50 nodes trying to transmit packets to the network. Its assumed that a maximum of 4 packets from the devices can fit into one RB. Further parameters are chosen as $p_1 = 0.1$, $p_2 = 0.1$, $n = 5$, $m = 5$. In the analytical results, there are some variations at points that the amount of the ceiling function is increased to a higher integer value. These variations become less as the aggregation period increases. The normalized required RBs function converges to the value $\frac{1}{\lceil \frac{s}{c} \rceil}$ regardless of traffic load. Note that if the $\frac{s}{c}$ is an integer value, then there in no gain in aggregating the data at the gateway in term of resource allocation.

C. Delay analysis

The packet aggregation scheme improves the efficiency of communications by adding extra delay for accumulating packets. When the aggregation method is not applied, the transmission latency consists of transmission delays between a device to the gateway, and from the gateway to the eNodeB. However, when the aggregation method is utilized, the aggregation period increases the transmission latency. In the both scenarios, packets might experience different latencies due to employed MAC protocol, packet loss and RA failure. Fig. 5 illustrates packet transition from a device to the eNodeB. This model can be used to analyze the average communication latency.

When a packet is generated, the device initially attempts to transmit the packet, if the packet is not delivered successfully after m transmission attempts, the packet will be dropped. The packet goes to the state $S_{1,i}, i \in \{1, \ldots, m\}$ when the device

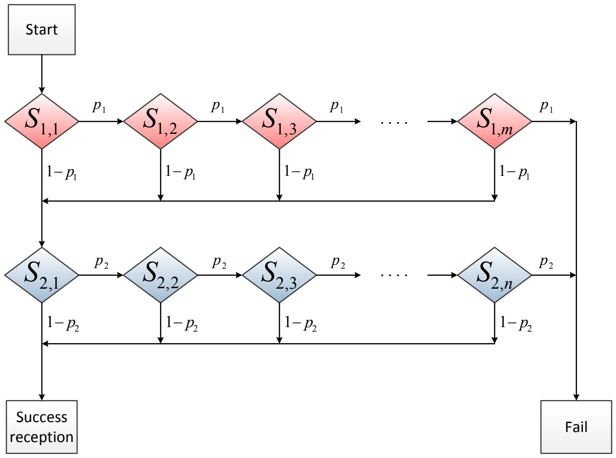

Fig. 5. Data flow for a generated packet.

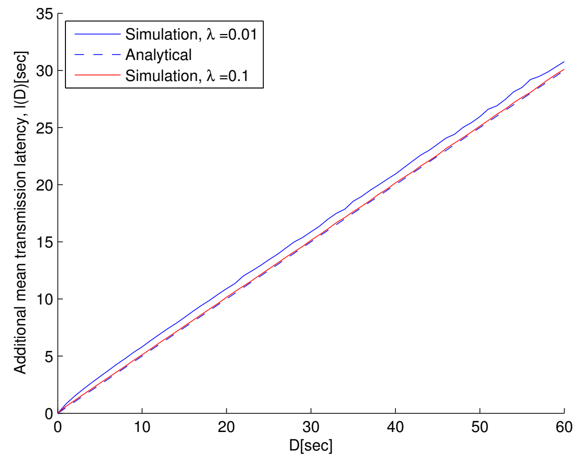

Fig. 6. Simulation and analytical results for additional mean transmission latency.

performs the ith attempt to send the packet to the gateway. After each attempt, the packet transmission might be successfully with probability $1-p_1$, or unsuccessful with probability p_1. Additionally, we can define the mean transmission latency $d_{1,i}$ as the mean time that takes a packet be delivered to the gateway after ith attempt from the time instant it has been generated.

The transmission latency between the gateway and the eNodeB consists of link establishment and data transmission delays. However, the link establishment delay is more dominant as the RA procedure is performed over a shared PRACH. When the gateway intends to deliver data to the eNodeB, it first needs to establish a link. It is assumed that the gateway can perform RA process successfully with probability of $1 - p_2$. If the RA process is failed, the gateway performs RA process again. The maximum number of RA process is limited to n times for each data transmission. It is also assumed that data is transmitted without failure when the link is established. In this model, state $S_{2,j}, j \in \{1, \ldots, n\}$ denotes jth attempt of the gateway to establish a link. Data is forwarded to the gateway only if the RA process is successful within n attempts, otherwise data is dropped. We also define the mean transmission latency $d_{2,j}$ as the mean time which data is delivered to the eNodeB after jth RA attempt.

In order to compare the communication latency for the packet aggregation scheme, we define an additional mean transmission latency. It is the difference between the mean transmission latency for the case when the aggregation method is applied and the case without the aggregation, that is:

$$l(D) = Mean\ transmission\ latency\ with\ aggregation - mean\ transmission\ latency\ without\ aggregation.$$

When the packet aggregation scheme is not applied, the mean transmission delay is calculated for packets which are delivered to the network successfully. This is achieved by going through all the combinations of delays related to both $S_{1,i}, i \in \{1, \ldots, n\}$ and $S_{2,j}, j \in \{1, \ldots, n\}$ states. The mean transmission latency $\bar{T}$ can be expressed as follows:

$$\bar{T} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} p_1^{i-1}(1-p_1)p_2^{j-1}(1-p_2)(d_{1,i}+d_{2,j})}{(1-p_1^{m-1})(1-p_2^{n-1})}.$$

When the aggregation method is utilized, received packets in the gateway experience additional delays which are due to the aggregation. It was mentioned that the aggregation period begins with the arrival of the first packet and ends after a period of D. If packets arrive randomly, the first packet will always suffer a delay of D at the gateway before the gateway begins the process of RA, while other packets arriving during the aggregation period will, in average, suffer a delay of $D/2$. Hence, the average aggregation delay for packets arriving during a single aggregation period is $\frac{D+N\lambda D(1-p_1^m)\cdot D/2}{1+N\lambda D(1-p_1^m)}$ which can be approximated by $D/2$ if the number of packets arrived during the aggregation period is high. The additional mean latency is the average of suffered delays due to the aggregation, i.e.:

$$l(D) = \frac{D + N\lambda D(1-p_1^m)\cdot D/2}{1+N\lambda D(1-p_1^m)} \approx D/2.$$

Fig. 5 shows the simulation and analytical results of additional mean latency when 50 nodes transmit packets to the network. It is assumed that $d_{1,i} = 10i$ ms and $d_{1,j} = 20j$ ms, $p_1 = 0.1$, $p_2 = 0.1$, $m = 5$ and $n = 5$. It is observed that the extra latency due to the aggregation has a linear relation to the aggregation period. Another observation is that the simulation results are closer to the analytical approximation for the higher traffic from nodes.

IV. Conclusion

Massive deployment of machines bring technical challenges concerning the RA and radio resource allocation. In this paper, we analyzed the performance of a packet aggregation scheme in capillary networks for delivering data from machine devices to the LTE network. The gateway aggregates packet during the aggregation period and then forward data to the eNodeB. Results showed that this scheme improves the communication efficiency by reducing RA interactions and the radio resources needed for data transmission. In addition, closed form formulas for RA process, radio resource allocation, and average latency as the function of the aggregation period were driven. These formulas can be used to find the optimum aggregation period to meet application and network constrains.

Acknowledgment

This research work has been partially supported by EIT ICT Labs under M2MRISE project. The authors would like to thank Pasi Lassila and Samuli Aalto for useful discussions.

References

[1] R. Ratasuk, S. Iraji, K. Hugl, L. Wang, and A. Ghosh, *"Performance of Low-Cost LTE Devices for Advanced Metering Infrastructure,"* IEEE VTC-Spring, vol. 1, no. 5, June 2013, pp 2-5.

[2] MZ. Shafiq et al., *"A first look at cellular machine-to-machine traffic: large scale measurement and characterization,"* ACM SIGMETRICS Performance Evaluation Review, 2012.

[3] 3GPP TS 22.368, *"Service requirements for Machine-Type Communications (MTC),"* v.13.1.0, 2014.

[4] A. Laya, L. Alonso, and J. Alonso-Zarate, *"Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives,"* .

[5] 3GPP TSG RAN WG2 #71 R2-104662, *"MTC simulation results with specific solutions,"* ZTE, Madrid, Spain, 23rd August 2010.

[6] X. Yang, A. Fapojuwo, and E. Egbogah *"Performance Analysis and Parameter Optimization of Random Access Backoff Algorithm in LTE,"* Proc. 2012 IEEE Vehicular Technology Conference (VTC Fall), 2012, pp. 15.

[7] J. Cheng, C. Lee, and T. Lin, *"Prioritized Random Access with dynamic access barring for RAN overload in 3GPP LTE-A networks,"* Proc. 2011 IEEE GLOBECOM Workshops (GC Wkshps), Dec. 2011, pp. 368372.

[8] K. Zhou and N. Nikaein, *"Packet aggregation for Machine Type Communications in LTE with Random Access Channel,"* Proc. IEEE WCNC, 2013.

[9] H. Shariatmadari et al., *"Delay Analysis of Network Architectures for Machine-to-Machine Communications in LTE System,"* Proc. IEEE ICT, May 2014.

[10] 3GPP TS 36.321, *"Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Medium Access Control (MAC) protocol specification,"* Release 10, 2011.